DOCUMENT RESUME

ED 185 960 HE 012 688

AUTHOR Hardesty, Larry
TITLE Use of Multiple Regression to Predict Academic Achievement at a Small Liberal Arts College.
PUB DATE 1 May 80
NOTE 32p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; Achievement Tests; *Aptitude Tests; Class Rank; *College Bound Students; *College Students; Correlation; Grade Point Average; Higher Education; *Multiple Regression Analysis; *Predictive Validity; Predictor Variables; Small Colleges
IDENTIFIERS *DePauw University IN

ABSTRACT

The relationship between academic success at DePauw University and such commonly used predictors as tested ability and academic success in high school was examined. The various subtleties of the multiple regression research method were also examined. Subjects were 1758 students who entered DePauw University during the fall semester of 1973, 1974, or 1975. Academic success was determined by the outgoing grade point average (OGPA) at the time the student left the school. Test scores included in the study were the verbal (SATVB) and the quantitative (SATQ) sections of the Scholastic Aptitude Tests. The scores from the battery of the College Entrance Examination Board achievement tests were also considered. Intercorrelations indicate that the scores for this study group are similarly related as they were for groups in previous research studies. High school rank tends to have the highest correlation with OGPA, and it is followed by SATVB. High school rank is relatively independent from the other scores, but the aptitude and achievement test scores are highly intercorrelated. Other correlations are anlayzed. It is concluded that the use of the multiple regression did result in a prediction equation that appears to be very useful in a pragmatic sense. Some problems with multiple regression are cited. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED185960

USE OF MULTIPLE REGRESSION TO PREDICT ACADEMIC ACHIEVEMENT AT A SMALL LIBERAL ARTS COLLEGE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Larry Hardesty

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U S DEPARTMENT OF HEALTH EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

by

Larry Hardesty
Head of the Reference Department
Roy O. West Library
DePauw University

May 1, 1980

HE 012 688

INTRODUCTION

Statement of Problem

> Concern with the prediction of academic performance has increased during recent years. One reason for this is the much-publicized growth in the student population. . . . for college admissions officers the selection of students is more difficult than ever before because the increase in the sheer number of applicants paralleled by growth in the number of highly qualified candidates. (Lavin, The Prediction of Academic Performance, 1965, p. 11)

Today many admissions officers probably look back and see the problem of the 1960's as a very pleasant problem they would still like to have. Today, however, there is increased competition among schools for students, and the prediction of academic performance has become more important in terms of predicting whether the marginal candidate will be turned away or not.

For example, at DePauw University, a recent projection forecasted approximately a twenty percent decline in the enrollment over the next ten years. Such small private schools are particularly caught between the problem of trying to maintain a tradition of quality and trying to maintain the type of student body, both in terms of quantity and quality, to support this tradition. Certainly the admissions officer wants to find as many highly qualified candidates as possible (and to predict their academic success as accurately as possible), but he also needs to recruit a sufficient number of students to maintain the student body. For many students it becomes important to be able to predict whether they will need extra guidance and remedial instruction in order to complete the instructional program.

Such predictors as high school grades and scores on standardized aptitude tests have long been used to aid in the prediction of academic success

in college. However, with the changing composition of the student body, the predictors and the weights given them in developing prediction equations used to predict the academic performance of a group of highly qualified students may not be adequate in predicting the academic performance of a group of students with a wider range of abilities.

The purpose of this study, therefore, is to examine the relationship between academic success at DePauw University and such commonly used predictors as tested ability and academic success in high school. The second purpose, is to explore the various subtleties of a statistical method (in this case, multiple regression).

Review of the Literature

There has been no shortage of investigations of the relationship between certain variables and academic achievement in college. Literally hundreds of studies have been completed, and there are now a number of reviews of the literature available (Giddens, 4-13). A majority of the studies have found the highest and most consistent correlation coefficients of college academic performance with some measure of academic performance in high school. In a review of 300 studies examined by Garret and the 194 examined by Emme, a majority of the studies concluded that high school performance was the best single predictor of college performance. Garret found the correlation coefficient between high school grades and college performance to be an average of +0.55 and to range from +0.18 to +0.83.

The wide variations in the range of the coefficients are probably due to the variations in the restrictions in range of students accepted according to rank at the various colleges and universities studied (Michael, 378). As Fricke points out, ". . . low correlations are to be desired, high correlation

will be obtained when no selection procedure is used or when an effective one is used" (Fricke, 43). The resulting correlations might be even higher if certain pre-college predictors were not used.

The second most commonly used predictors in addition to high school achievement are aptitude test scores. Perhaps the most widely used tests are the College Board Scholastic Aptitude Tests. Philip DuBois, writing in the Seventh Mental Measurements Yearbook, observed "on no other test have there been expended so many man-years of planning, item-writing, statistical analyses, and evaluation in its various forms" (DuBois, 646). He also noted that the Scholastic Aptitude Tests (SAT) have reasonably good validations for predicting college achievement. DuBois concluded:

> For schools of engineering, science, business, education, and liberal arts, the high school record is more valid than the SAT, but the SAT supplements the high school record with additional valid variance. Except in engineering, the SAT-V is generally more predictive than the SAT-M (DuBois, 647).

Chase reported that typical correlations of the SAT-V score with college grade-point average are .33 for men and .41 for women; the SAT-M score correlations with a grade-point average are .30 for men and .36 for women. For the same group, high school record correlates with college grades at .47 for men and .54 for women (Chase, 180). DuBois reported that multiple correlations using SAT-V, SAT-M, and high school record with colleges grades are in the order of .62 (DuBois, 647). Various batteries of tests with high school rank have multiple correlations with college grades from .60 to .70 (Lavin, 52).

Statistical Methods Used

This study did not attempt to move away from a long standing pattern of multiple regression as the technique used to determine predictors of college academic achievement. Multiple regression is often used:

> (1) to find the best linear prediction equation and evaluate its prediction accuracy; (2) to control for other confounding factors in order to evaluate the contribution of a specific variable or set of variables; (3) to find structural relations and provide explanations for seemingly complex multivariate relationships (SPSS Manual, 321).

Certainly the purpose of this study was to develop an easily applied formula for predicting college academic performance. Part of the ease of application included a reduction in the number of predictors involved and an accounting for the relationship between certain commonly used predictors. For example, verbal skill may be an important part of the resulting scores from various tests, including the SAT-M test. Chase noted that the overlap between the two scores of usually between .65 and .70 (Chase, 179). This intercorrelation score has been increasing since the 1950's (Wallace, 649).

The particular multiple regression procedure used was that available through the SPSS program at Indiana University. The stepwise procedure was used with deletion of subjects with missing scores. According to Herzberg, the stepwise procedure is an excellent method of examining the subset of predictors to determine their contribution to the prediction. Through this method the number of predictors used can be narrowed if some of them do not add significantly to the prediction equation (Herzberg, 10).

As already indicated, there are no shortage of studies using this method in predicting college academic performance. The procedure is common place at many schools, and the studies reported in the recent literature tends to explore the more subtle aspects of the technique. Related studies at similar schools to DePauw University have been published by Michael and Jones, Vick and Hornaday, and Mann.

PROCEDURES

Description of Sample

The study is limited to 1758 students who entered DePauw University during the fall semester of 1973, 1974, or 1975. Not included were transfer students, students who did not provide SAT scores (approximately 160), and students who did not complete the first academic year (approximately 75).

Academic success was determined by the outgoing grade point average (OGPA) at the time the student left the school. For the entering classes of 1973 and 1974, OGPA could be the score at the time of graduation, transferring, or dropping out of college. For the entering class of 1975, OGPA is the score at the time of graduation, transferring, dropping out of the college, or (for the bulk of this class) at the end of the seventh semester. The OGPA is based on a four point scale (4=A; 3=B; 2=C; 1=D; and 0=F).

The test scores included in this study are the verbal (SATVB) and the quantitative (SATQ) sections of the Scholastic Aptitude Tests. These scores were combined (SATTOT) for further analysis. Also, included are the scores from the battery of the College Entrance Examination Board (CEEB) achievement tests. These tests include a vocabulary test (VOCA), a level of comprehension test (LEV), and a speed of comprehension test (SPD). The score for the English composition test included in this battery was not used due to a computer error.

The achievement test scores are reported in raw score form. The SAT tests are reported in the usual manner using the standard scoring system where average performance is considered to be 500 and the standard deviation to be 100 (Chase, 179). High school rank (HSR) is reported as a percentile, i.e. for someone ranking number one in their class they would be reported as having a rank of .99. This method was used to ensure a positive correlation between OGPA and HSR (assuming that the relationship was significantly positive).

Other available variables included were: (1) declared major field of each student; and (2) whether the students graduated or not (for the 1973 and 1974 classes.)

The information was obtained through the cooperation the registrar's office of DePauw University. Information such as the SAT scores and the achievement test scores were taken from available records through the campus computer center and punched by machine onto a deck of cards. Other information, such as the high school rank (which had to be computed manually for each student) and the OGPA were taken from the individual student's file in the registrar's office, and keypunched manually onto the deck of cards. At all times the confidentiality of individual student records was maintained.

Logical Structure and Assumptions

It is assumed that college grades are an adequate measure of academic success, although it is noted that "college grades are a complex, heterogeneous criterion, influenced by variables other than academic ability" (Chase, 180). Also, much the same can be said for high school grades as a measure of academic success in high school (Lavin, 51). It is also noted that grades are not the only criterion to be used in determining the potential contribution an individual might make to society (Lavin, 16). Other factors, such as creativity, are often not reflected through grades. As noted in the review of the literature, SAT tests, while they are assumed for purposes of this study to be an adequate measure of aptitude, are not without their limitations.

Concerning the selection of students, it was assumed that one semester's grade point average (or less than one semester's grade point average) was less valid as a criterion than one academic year's grade point average or more (although a number of studies use the first semester or the freshman year grade point average). The students excluded because of lack of SAT scores are

assumed not to be a special group that would skew the results of this study (and this assumption is shared by the director of DePauw University's Testing and Research Bureau).

RESULTS

The intercorrelations (as displayed in Table One) indicate that the scores for this group are similarly related as they were for groups in previous research studies. HSR rank tends to have the highest correlation with OGPA, and it is followed by SATVB. HSR rank is relatively independent from the other scores (with the .32 to .39 rank), but the aptitude and achievement test scores are highly intercorrelated. This problem of intercorrelation among independent variables is known as multicollinearity, and common solution is to merely discard one of the variables (Hamburg, 515). It should be noted that VOCA, LEV, and SPD all have an intercorrelation with SATVB at .68 and above. Even SATVB and SATQ intercorrelate at .56. At this point it would be a safe assumption that the VOCA, LEV, and SPD scores will not add significantly to a prediction equation once the variance accounted for by SATVB is included. The results of a stepwise multiple regression procedure displayed by Table Two confirm this assumption.

Although the various achievement tests did come into the equation (only to the fourth step is displayed), they did not add significantly to the multiple correlation. This high correlation between two or more independent variables is termed by some authors as redundancy, and the number of redundant independent variables is termed reptitiveness (Gordon, 596). This is often a problem when selecting significant variables, and the stepwise regression procedure allows the research to examine the contribution of each variable.

As redundant independent variables are introduced into a regression problem, the common predictive value is averaged over all their regression

TABLE ONE

PREDICTOR INTERCORRELATIONS, MEANS, AND STANDARD DEVIATIONS

FOR ALL ENTERING STUDENTS, 1973, 1974, 1975*

Predictors	Means	Standard Deviations	Predictors						Criterion
			HSR	SATVB	SATQ	VOCA	LEV	SPD	OGPA
HSR	.77	.22		.38	.39	.32	.32	.32	.58
SATVB	499.50	90.60			.56	.76	.68	.76	.44
SATQ	540.97	94.20				.42	.47	.48	.41
VOCA	41.50	7.60					.63	.67	.35
LEV	21.50	4.80						.82	.36
SPD	33.90	10.90							.36
Criterion									
OGPA	2.81	.60							

*Includes only students who completed the entire battery of test, i.e. 1587 out of 1758.

TABLE TWO

STEPWISE MULTIPLE REGRESSION DISPLAY

USING ALL VARIABLES AND ALL STUDENTS*

	Variable	b-weight	constant	significance	R
Step One					
	HSR	1.5781	1.59	.001	.57895
	MR = 1.5781(HSR) + 1.59				
Step Two					
	HSR	1.3164		.001	
	SATVB	.0017	.94	.001	.62528
	MR = 1.3164(HSR) + .0017(SATVB) + .94				
Step Three					
	HSR	1.2515		.001	
	SATVB	.0013		.001	
	SATQ	.0007	.79	.001	.63190
	MR = 1.2515(HSR) + .0013(SATVB) + .0007(SATQ) + .79				
Step Four					
	HSR	1.2470		.001	
	SATVB	.0011		.001	
	SATQ	.00069		.001	
	LEV	.00574	.78	.092**	.63275
	MR = 1.2470(HSR) + .0011(SATVB) + .00069(SATQ) + .00574(LEV) + .79				

*Included only those students who completed the entire battery of tests, i.e. 1587 out of 1758.

**Contribution of variables after the third step no longer significant at .05 level.

coefficients. A result is that all their regression coefficients decline in absolute value. Gordon has discussed this situation in some detail (Gordon, 595). A common error as a result of including several highly intercorrelated variables is to underestimate the contribution of a particular variable, i.e. verbal ability, which may be a significant part of several of the variables.

Table Two indicates that the achievement test scores can be excluded since they do not add significantly to the prediction equation. This also works well in a practical sense because the SAT and HSR scores are usually available before a student is accepted into college, but the achievement test scores are the result of tests after the student has been accepted. Even if one of the achievement test had come into the equation earlier, it may have been necessary to force the SAT scores in before it.

While adding predictors to a regression equation may increase the multiple correlations, the greater the number of predictors, the more unstable may be the sample regression weights. A small number of predictors may do equally as well (as indicated in this case), and the weights for a subset of predictors would be more stable in future samples due to the smaller n (Herzberg,9).

Table Three displays what can happen when a high degree of intercorrelation is present. Since the SATVB and SATQ tests are added to determine the SATTOT score, they intercorrelate with it to a high degree (.88 and .89 respectively). When shown in a stepwise procedure all three variables do come into the formula, but in this case SATTOT comes in as a negative b-weight after SATQ is introduced (as indicated in Table Four). While this may be statistically sound, it may take considerable explanation to convince the statistically unsophisticated of the logic of this equation. (SATTOT is not to be confused with a supressor variable since it does correlate significantly

TABLE THREE

PREDICTOR INTERCORRELATIONS, MEANS, AND STANDARD DEVIATIONS

FOR SELECTED VARIABLES

FOR ALL ENTERING STUDENTS, 1973, 1974, 1975*

Predictors	Means	Standard Deviations	Predictors				Criterion
			HSR	SATTOT	SATVB	SATQ	OGPA
HSR	.769	.22		.43	.38	.39	.57
SATTOT	1036.36	164.81			.88	.89	.46
SATVB	497.28	90.66				.57	.43
SATQ	539.57	94.96					.40
Criterion							
OGPA	2.80	.60					

*Includes 1758 students who completed SAT tests

TABLE FOUR

STEPWISE MULTIPLE REGRESSION DISPLAY

FOR SELECTED VARIABLES

FOR ALL ENTERING STUDENTS, 1973, 1974, 1975*

	variables	b-weight	constant	significance	R
Step One					
	HSR	1.569	1.594	.001	.57273
Step Two					
	HSR	1.248		.001	
	SATTOT	.000997	.807	.001	.62231
Step Three					
	HSR	1.248		.001	
	SATTOT	.000664		.001	
	SATVB	.000687	.810	.009	.62422
Step Four					
	HSR	1.243		.001	
	SATTOT	-.00191		.052	
	SATVB	.00324		.001	
	SATQ	.00262	.797	.008	.62617

*Includes 1758 students who completed SAT tests

(.46) with the criterion OGPA (see Table Three). In any case, taking SATTOT out of the procedure involves little, if any, loss in prediction (from .626 to .625, see Tables Four and Five).

The SAT scores can be combined in several ways. Table Six displays the results of efforts to use various combinations of the SAT scores and a cross-validation (actually a double cross-validation) of the resulting equations. Use of the SATVB score alone with HSR results in almost the same multiple correlation as use of either SATTOT or SATVB and SATQ with HSR. This reflects that SATQ makes little additional contribution (although statistically significant) to predicting OGPA.

The correlations between the actual OGPA for a particular year and the predicted OGPA (generated through use of equation developed on a different year) hold up reasonably well. A large drop is often the result from year to year when the predictive variables are used for selection purposes (Vick and Hornaday, 798), but this does not appear to be the case. There is some change from 1973 to 1974 in terms of predictability of each class, and the cause of this is not entirely clear. There may be less selectivity involved for the 1974 class, and, as a result, more range (as noted in Table Seven in the standard deviations of the SATVB and SATQ scores, but this increase in standard deviation is not reflected in the HSR scores).

There are also differences in the weights given the predictor variables from year to year. From 1973 to 1974 more weight is given the HSR and less to the SAT scores (see Table Six). From an examination of the 1975 group's means and standard deviations for each variable (not displayed), it was determined that the 1974 group was more similar to the 1975 group, and a cross validation using the 1973 equations and the 1974 equations indicated that the 1974 equations better predicted, i.e. correlated with, the 1975 OGPA (results are not displayed).

TABLE FIVE

STEPWISE MULTIPLE REGRESSION DISPLAY

USING HSR, SATVB, AND SATQ TO PREDICT OGPA

FOR ALL ENTERING STUDENTS, 1973, 1974, 1975

	variables	b-weight	constant	significant	R
Step One					
	HSR	1.569	1.594	.001	.57273
Step Two	HSR	1.306		.001	
	SATVB	.001697	.9531	.001	.61862
Step Three					
	HSR	1.243		.001	
	SATVB	.001325		.001	
	SATQ	.000721	.7974	.001	.62512

MR = 1.243(HSR) + .001325(SATVB) + .000721(SATQ) + .7974

*(n = 1758)

TABLE SIX

PREDICTION EQUATIONS, MULTIPLE CORRELATIONS, AND

CROSS-VALIDATION CORRELATIONS OF ACTUAL AND PREDICTED GRADES

Predictors	R for 1973 (N=620)	R for 1974 (N=595)	1973 to 1974	1974 to 1973
HSR and SATTOT	.59299	.63677	.6187	.5798
HSR and SATVB	.58207	.64166	.6320	.5753
HSR, SATVB, and SATQ	.59513	.64287	.6231	.5809

Prediction Equations

HSR and SATTOT

1973 MR = 1.08(HSR) + .00011(SATTOT) + .709

1974 MR = 1.57(HSR) + .00076(SATTOT) + .745

HSR and SATVB

1973 MR = 1.16(HSR) + .0019(SATVB) + .92

1974 MR = 1.58(HSR) + .0014(SATVB) + .80

HSR, SATVB, and SATQ

1973 MR = 1.077(HSR) + .00013(SATVB) + .00107(SATQ) + .68

1974 MR = 1.55(HSR) + .0013(SATVB) + .0003(SATQ) + .73

Actual and Predicted Scores Using HSR, SATVB, and SATQ

	Mean	S.D.	r
1973 Actual Scores	2.7957	.6218	
1973 Predicted Scores Using 1974 formula	2.7627	.4230	.5809
1974 Actual Scores	2.7730	.6137	.6231
1974 Predicted Scores Using 1973 formula	2.7934	.3574	

TABLE SEVEN

DESCRIPTIVE INFORMATION

BY ENTERING CLASS

Class	Predictor	Mean	Standard Deviation
1973	HSR	.7667	.2239
(N=620)	SATVB	505.7258	85.7811
	SATQ	545.3387	89.6668
	Criterion		
	OGPA	2.7957	.6218
1974	HSR	.7839	.2039
	SATVB	494.2017	92.8092
	SATQ	540.6605	97.9300
	Criterion		
	OGPA	2.7730	.6100

It can be speculated that this is because of a trend toward less selectivity due to a decrease in the number of applicants (beginning in 1974).

An attempt was made to moderate on HSR to determine if the predictive power of SATVB and SATQ could be improved. A moderator is defined as "variables, quantitative or qualitative, which improve the usefulness of a predictor (or predictors) by isolating subgroups of individuals for whom a predictor or a set of regression weights are especially appropriate" (Banas, 17, as quoted by Wood, 311). In this case an attempt was made to tap some of the underlying variables, such as motivation, that determine HSR. It can be speculated that students who are in the upper sections of their HSR may be more motivated to "live-up" to their aptitude (as determined by SAT scores), and, thus, the SAT scores (particularly the SATVB score) may be more predictive of OGPA as HSR increases.

For the entire group (n=1758), SATVB correlates with OGPA at .43 and SATQ correlates with OGPA at .40 (see last section of Table Eight). When the group is equally divided on the basis of HSR, the correlations drop for both groups, with the correlations and the multiple correlations being larger in the upper half according to HSR (see Table Eight). This would tend to confirm this theory. However, this increase does not continue as the top half is further subdivided (which might be expected due to restriction in range). Some accounting for this restriction for range may be needed to better explore the relationships examined here. Such a restriction in range may account for why predictability of performance for graduate school is usually lower than for the college level. The graduate students are a more select group with a considerably narrower range of abilities (of higher quality) (Lavin, 51).

Other researchers have generally found females to be more predictable than males in academic performance (Lavin, 52). Using only HSR and SATTOT as

TABLE EIGHT

MEANS, STANDARD DEVIATIONS, INTERCORRELATIONS,

AND MULTIPLE CORRELATION COEFFICIENTS USING

SATVB AND SATQ FOR PREDICTING OGPA

IN SELECTED HSR GROUPS

Group	Predictor	Means	Standard Deviations	Predictors		Criterion
				SATVB	SATQ	OGPA
Bottom Half HSR 1973-1975 (N=878)	SATVB	462	81		.47	.26
	SATQ	500	87			.19
	Criterion					
	OGPA	2.47	.54			

R = .279 MR = .0015(SATVB) + .0005(SATQ) + 1.49

Top Half HSR 1973-1975 (N=881)	SATVB	531	86		.51	.32
	SATQ	578	80			.29
	Criterion					
	OGPA	3.12	.49			

R = .36 MR = .0013(SATVB) + .00097(SATQ) + 1.84

Above .90 HSR 1973-1975 (N=712)	SATVB	541	83		.46	.30
	SATQ	585	83			.27
	Criterion					
	OGPA	3.20	.45			

R = .334 MR = .00118(SATVB) + .000907(SATQ) + 2.03

TABLE EIGHT (cont.)

Group	Predictor	Means	Standard Deviations	Predictors SATVB	Predictors SATQ	Criterion OGPA
Above .95 HSR 1973-1975 (N=439)	SATVB	556	79		.45	.31
	SATQ	605	78			.23
	Criterion					
	OGPA	3.30	.39			

R = .332 MR = .0013(SATVB) + .00054(SATQ) + 2.28

Above .98 HSR 1973-1975 (N=233)	SATVB	565	78		.40	.31
	SATQ	620	74			.19
	Criterion					
	OGPA	3.40	.38			

R = .31 MR = .0013(SATVB) + .0004(SATQ) + 2.365

Correlation Coefficients for Entire Group (N=1738)

	SATVB	SATQ	OGPA
SATVB		.57	.43
SATQ			.40

predictors, this study would tend to reverse this conclusion, although the differences are small (see Table Nine). The females tend to have a smaller standard deviation in their HSR (and a higher HSR), and since this predictor has the largest weight, this may explain why the females in this study are less predictable.

When the ranges are relatively unrestricted, one can examine the role of the various predictors in predicting the OGPA of the students within the various disciplines or subject areas. The results of such an examination are displayed by Table Ten. As one might expect, the SATQ score has a larger b-weight than the SATVB score for science majors. In fact, the SATVB score does not make a significant contribution.

This role is reversed for the humanities majors and social science majors (although the SATQ does make a significant contribution in the case of the humanities majors). The various relationships are less clear in terms of the education majors. For them, the SATVB score does not add significantly to the predictability, which is not what would be reasonably expected. The only significant difference from the previous three groups mentioned that can be determined through visually scanning the information is in the size of the standard deviation of the OGPA and the size of the SATVB and SATQ scores. Also, it should be noted that the R is smaller.

The explanation for the nursing students is more evident (although initially one has to question the complete reversal of the predictor in the equation). The selection into the nursing program is more competitive than selection into the other programs, and it would appear that some of the selection is based on the HSR of the students. With a truncation in the range of a predictor variable, and a result a reduction in variance, we can expect a decrease in the size of the predictive coefficients. This appears to be the

TABLE NINE

MEANS, STANDARD DEVIATIONS, INTERCORRELATIONS,

AND MULTIPLE CORRELATION COEFFICIENTS USING

HSR AND SATTOT FOR PREDICTIONG OGPA

FOR MALES AND FEMALES

Group	Predictor	Means	Standard Deviations	Predictors		Criterion
				HSR	SATTOT	OGPA
Males						
	HSR	.722	.24		.47	.589
	SATTOT	1054	158			.46
	Criterion					
	OGPA					
		2.75	.64			
	R = .62	MR = 1.27(HSR) + .000989(SATTOT) + .78				
Females						
	HSR	.81	.18		.49	.54
	SATTOT	1018	168			.50
	Criterion					
	OGPA	2.80	.56			
	R = .608	MR = 1.15(HSR) + .00105(SATTOT) + .84				

TABLE TEN

MEANS, STANDARD DEVIATIONS, INTERCORRELATIONS

AND MULTIPLE CORRELATION COEFFICIENTS USING

HSR, SATVB, AND SATQ TO PREDICT OGPA

FOR STUDENTS BY MAJOR

Group	Predictors	Means	Standard Deviations	Predictors			Criterion
				HSR	SATVB	SATQ	OGPA
Science Majors	HSR	.86	.16		.32	.28	.59
	SATVB	522	90			.55	.39
(N=260)	SATQ	603	86				.43
	Criterion						
	OGPA	3.11	.54				

R = .66 MR = 1.606(HSR) + .0015(SATQ) + .00058(SATVB)* + .495

Humanities Majors	HSR	.78	.21		.41	.41	.58
	SATVB	510	91			.58	.49
(N=333)	SATQ	526	94				.41
	Criterion						
	OGPA	2.88	.51				

R = .65 MR = 1.09(HSR) + .0014(SATVB) + .000424(SATQ)* + 1.05

TABLE TEN (cont.)

Group	Predictor	Means	Standard Deviations	Predictors			Criterion
				HSR	SATVB	SATQ	OGPA
Social Science Majors	HSR	.75	.23		.42	.42	.60
	SATVB	502	88			.51	.50
(N=545)	SATQ	544	88				.43
	Criterion						
	OGPA	2.86	.55				

R = .67 MR = 1.09(HSR) + .0016(SATVB) + .00071(SATQ) + .845

Education Majors	HSR	.75	.22	.33	.38	.52
	SATVB	440	89		.60	.38
(N=144)	SATQ	483	95			.47
	Criterion					
	OGPA	2.87	.45			

R = .60 MR = .809(HSR) + .0012(SATQ) + .00046(SATVB)* + 1.446

Nursing Majors	HSR	.84	.18	.18	.34	.41
	SATVB	483	72		.45	.46
(N=88)	SATQ	513	81			.47
	Criterion					
	OGPA	2.80	.47			

R = .62 MR = .00167(SATQ) + .00189(SATVP) + .687(HSR) + .453

*Predictor does not contribute at .05 level of significance

case with HSR with the nursing students, although the SAT scores may not be used for selection purposes since the correlation remains relatively high (.62).

Multiple regression is perhaps most appropriate for examination of the relationships between variables that are on a continuum. It is useful for predicting how well one will do (i.e. range of OGPA), but it is a little less useful in predicting if a student will graduate or not (which is a dichotomous variable). Sometimes an admissions officer may be more interested in whether a student will complete four years of college (i.e. graduate) or not rather than how well the student will do. Discriminate analysis may be a more useful technique when attempting to determine membership in two or more groups. A question in this study was whether the variables available would serve as an adequate set of discriminating variables to distinguish graduates from those students who could not graduate due to low OGPA (below 2.00).

Initially all the variables were included in the analysis, including the achievement test scores, and all the students were included from the 1973 and 1974 classes (and those members of the 1975 class who had graduates). Only approximately twelve percent of this group had not graduated because of low OGPA. When using the discriminate analysis program in SPSS, priorities in terms of probabilities of group membership can be established. However, when this was done, the number of students to graduate was over predicted because of the large percent in this group.

A second study was completed attempting to establish two equal groups. Group one consisted of ninety-four students who had not graduated and had an OGPA less than 2.00. The second group consisted of eighty-one students who had graduated with a OGPA between 2.00 and 2.40. It is probably with this group that an admission officers would have difficulty in determining who to admit to college or not to admit (hoping, of course, to admit those students who had a

reasonable chance of graduating.

A summary of the results of the use of this procedure are displayed by Table Eleven. Again, all the available variables were included. However, only the SATQ and HSR variables were entered into the equation under the default option and use of the stepwise method. The Wilks Lambda serves as an inverse measure of the discriminating power of the original variables which has not yet been removed by the discriminant functions. The larger the lambda is, the less information remaining. As indicated in Table Eleven, the variables included had little discriminating power in determining membership in the two groups. The eigenvalue is a measure in this case of the relative importance of this function, and it is very small in this case. Finally, the prediction results displayed by the discriminate analysis program of SPSS indicate the percent of known cases correctly classified (54.3 percent), and the result is little different than one would expect from chance (and probably not significantly different than chance).

CONCLUSIONS

Use of the multiple regression in this study does result in a prediction equation that appears to be very useful in a pragmatic sense. The objective here is the elimination of superfluous variables rather than to test theoretical hypotheses, and multiple regression works well for this with little additional manipulation or interpretation. With a large sample size, the intercorrelation among variables tends not to be a problem, particularly since this is not an attempt to force in certain variables to examine their relationship with the criterion more closely.

Some caution must be made in terms of use of this method over a period of time and the interpretation of results in relation to particular subgroups, i.e., the nursing students. Without realizing the results of restrictions of

TABLE ELEVEN

DISCRIMINATE ANALYSIS RESULTS

Predictors	Group One OGPA less than 2.00 and not graduate (N=94) Means	Standard Deviations	Group Two OGPA between 2.00 and 2.40 and graduated (N=81) Means	Standard Deviations
HSR	.54	.23	.58	.23
SATVB	449	80	441	80
SATQ	497	93	478	85
VOCA	37	8	37	8
LEV	19.2	5.1	18.4	4.9
SPD	29.0	11	27.7	9

Variables Entered	Wilks Lambda	Significant	Eigenvalue
SATQ	.98866	.161	
HSR	.97911	.163	.02133

Centroids of Group in Reduced Space

Group 1 .13481
Group 2 -.15645

Prediction Results	N of cases	Predicted Group 1	Predicted Group 2
Group 1	94	54 (53.2%)	44 (46.8%)
Group 2	81	36 (44.4%)	45 (55.6%)

54.3 percent of known cases correctly classified

academic achievement of nurses (or the same conclusion for HSR). Actually, it may be very important, and it may have been used as a major criterion for selection into the program. The paradox of the multiple regression technique is that once a significant predictive variable is identified and used for selection purposes, its weight decreases as new equations are developed using the students selected. This, of course, is an example of the "bouncing beta".

A second problem with multiple regression not previously mentioned is related to the inherit nature of the regression technique. The predicted scores tend to regress to the mean (hence the name of the technique). This phenomenon is indicated by the results at the bottom of Table Six where the actual OGPA has a larger standard deviation than the predicted OGPA generated using the multiple regression equation. The practical result of this for selection purposes is that OGPA of students with low OGPA scores are over predicted and the OGPA scores of students with high OGPA scores are under predicted.

For the low OGPA students this may be a desirable effect for selection purposes. From a visual examination of the predicted and actual scores, it is evident that very seldom did a student have an actual score above 2.00 (minimum needed for graduation) if the student was predicted to have less than a 2.00 OGPA. Thus, if a student was predicted to have less than a 2.00 OGPA, an admissions officer could reject the application of this student with a high degree of confidence in that decision.

However, at the other end of the continuum, very few students were predicted to have more than a 3.40 OGPA when actually a number of students did have a high OGPA. Use of multiple regression is less satisfactory for use in selecting students for an honors program.

Discriminate analysis did not prove to be a satisfactory procedure to predicting whether a marginal student would graduate or not. This is probably

understandable considering that from the means and standard deviations displayed by Table Eleven indicate that two groups used are almost indistinguishable using the variables in this study. There may be some satisfaction in knowing that the variables included cannot be used to make such fine distinctions through use of discriminate analysis. Perhaps other factors, such as personality characteristics and biographical information, needed to be included in order make valid discriminations.

BIBLIOGRAPHY

Chase, Clinton I. Measurement for Educational Evaluation. Reading Massachusetts: Addison-Wesley Publishing Company, 1974.

Darlington, Richard B. "Multiple Regression in Psychological Research and Practice," Psychological Bulletin, Vol. 69, No. 3 (1968), 161-182.

DuBois, Philip H. "College Board Scholastic Aptitude Test," The Seventh Mental Measurement Yearbook, ed. O.K. Buros (Highland, New Jersey: The Gryphon Press, 1972), 646-648.

Fricke, Benno G. "Prediction, Selection, Mortality, and Quality Control," College and University, Vol. 32 (Fall, 1956), 34-52.

Garrett, Hurley F. "A Review of Interpretation of Investigations of Factors Related to Scholastic Success in Colleges of Arts and Sciences and Teachers Colleges," Journal of Experimental Education, Vol. 18 (December, 1949), 91-138.

Giddens, Thomas R. "Disparity and Combined Score Patterns Exhibited on the SAT and Their Relationship to Academic Success at DePauw University." Unpublished Master's thesis, Indiana University, 1965.

Gordon, R.A. "Issues in Multiple Regression," American Journal of Sociology, Vol. 73 (1969), 592-616.

Hamburg, Morris. Statistical Analysis for Decision Making. New York: Harcourt, Brace and World, Inc., 1970.

Herzberg, Paul Albin. "The Parameters of Cross-Validation." Unpublished PhD dissertation, University of Illinois, 1967.

Lavin, David E. The Prediction of Academic Performance. New York: Russell Sage Foundation, 1965.

Mann, Sister M. Jacinta, "The Prediction of Achievement in a Liberal Arts College," Educational and Psychological Measurement, Vol. XXI, No. 2 (1961), 481-483.

Michael, William B. and Robert A. Jones. "Stability of Predictive Validities of High School Grades and of Scores on the Scholastic Aptitude Test of College Entrance Examination for Liberal Arts Students," Educational and Psychological Measurement, Vol. XXIII, No. 2 (1963), 375-378.

Nie, Morman H. and others. Statistical Package for the Social Sciences. 2nd ed. New York: McGraw-Hill Book Company, 1975.

Vick, Mary Catharine and John A. Hornaday. "Predicting Grade Point Average at a Small Southern College," Educational and Psychological Measurement, Vol. XXII, No.4 (1962), 795-799.

Wallace, Wimburn L. "College Board Scholastic Aptitude Test," The Seventh Mental Measurement Yearbook, ed. O.K. Buros (Highland, New Jersey: The Gryphon Press, 1972), 648-650.

Wood Donald A. and Michael J. Langevin. "Moderating the Prediction of Grades in Freshman Engineering," Journal of Educational Measurement, Vol. 9, No. 4 (Winter, 1972), 311-319.